

Relative 3D Positioning and 3D Convex Hull Computation from a Weakly Calibrated Stereo Pair

Luc Robert Olivier Faugeras

N° 2349

Septembre 1994

PROGRAMME 4

Robotique,
image
et vision

 *apport
de recherche*

1994

Relative 3D Positioning and 3D Convex Hull Computation from a Weakly Calibrated Stereo Pair

Luc Robert* Olivier Faugeras**

Programme 4 — Robotique, image et vision
Projet Robotvis

Rapport de recherche n°2349 — Septembre 1994 — 20 pages

Abstract: Assuming that we only know the epipolar geometry of a pair of stereo images, encoded in the so-called *fundamental matrix*, we show that some useful and intuitive three-dimensional information, such as relative positions of points and planes and 3D convex hulls, can be computed in the images without performing any three-dimensional reconstruction. We introduce the notion of *visibility*, which allows deriving those properties. Results on real data are shown.

Key-words: projective calibration, convex hull, fundamental matrix

(Résumé : *tsvp*)

*Email : lucr@sophia.inria.fr

**Email : faugeras@sophia.inria.fr

Positions Relatives de Points et de Plans et Calcul d'Enveloppes Convexes 3D à partir d'Images Faiblement Calibrées

Résumé : Dans ce rapport, nous considérons une paire d'images stéréo dont seule la géométrie épipolaire est connue, représentée par la *matrice fondamentale* du système de caméras. Nous montrons qu'il est possible de déterminer directement "dans les images", c'est-à-dire sans effectuer de reconstruction tridimensionnelle, des propriétés tridimensionnelles importantes et intuitives comme la position relative de points et de plans ou l'enveloppe convexe d'un ensemble de points. Nous introduisons la notion de *visibilité*, qui permet d'obtenir ces propriétés. Nous présentons des résultats obtenus sur des données réelles.

Mots-clé : calibration projective, enveloppe convexe, matrice fondamentale

1 Introduction

Recently, there has been a growing interest in trying to build “weak” descriptions of 3D space, that do not require camera calibration, but are nonetheless sufficient for performing sophisticated tasks like object recognition. The key point of these methods is the use of projective or affine invariants [7]. Affine reconstruction in the case of parallel projection has been presented by Koenderink et al. [4], and the projective case has been addressed in [1, 6, 10].

These approaches make use of some geometric relationships between the different images, that can be encoded in the 3×3 *fundamental matrix*. This matrix, which may be interpreted as a generalization of Longuet-Higgins essential matrix, can be directly computed from point correspondences in real world images [8, 2]. It contains the information about the epipolar geometry of the stereo rig.

Once the fundamental matrix has been computed for a system of cameras, one can build a three-dimensional description of any object for which correspondences have been established in two images, and then compute three-dimensional characteristics of this object.

The originality of our approach with respect to this somewhat classical scheme is that we do not use any intermediate three-dimensional description in order to derive some intuitive three-dimensional properties of the observed scene: all of the computation is done in the images. The three-dimensional information that we recover is twofold: First, we can determine the relative positions of two points in space with respect to a plane. Second, we apply this result to the computation of the convex hull of a set of points in space. After a brief characterization of the “weak-calibration” assumption, and the basic image properties of planar structures, we give the details of our approach. We conclude with some experimental results on convex hull computation, obtained on real data.

2 Weak calibration: the fundamental matrix

Let us consider a binocular stereo rig composed of two pinhole cameras 1,2. All the entities specific to a given camera are indexed with its number. C denotes the optical center, \mathcal{P}^r the plane of the retina, \mathcal{P}^o the optical plane (i.e. the plane that is parallel to the retina and passes through the optical center). The cameras are in general position. The transformation which associates to a point m_1 of image 1 its

epipolar line $\langle ep_2(m_1) \rangle$ in image 2 is a correlation of the projective plane \mathcal{P}^2 (i.e. a projective linear transformation that maps points into lines), whose matrix is called the *fundamental matrix*, denoted by \mathbf{F} [2]. \mathbf{F} has rank 2, its kernel is the 3-vector representing in homogeneous coordinates the epipole e_1 . All those notations are summarized in Figure 1. If points m_1, m_2 represent the same space point M , then their projective coordinate vectors (3-component vectors defined up to a scale factor, and denoted by bold-face characters) satisfy the Longuet-Higgins equation [5]

$$\mathbf{m}_2^t \mathbf{F} \mathbf{m}_1 = 0 \quad (1)$$

As a consequence, the fundamental matrix from image 2 to 1 is \mathbf{F}^t .

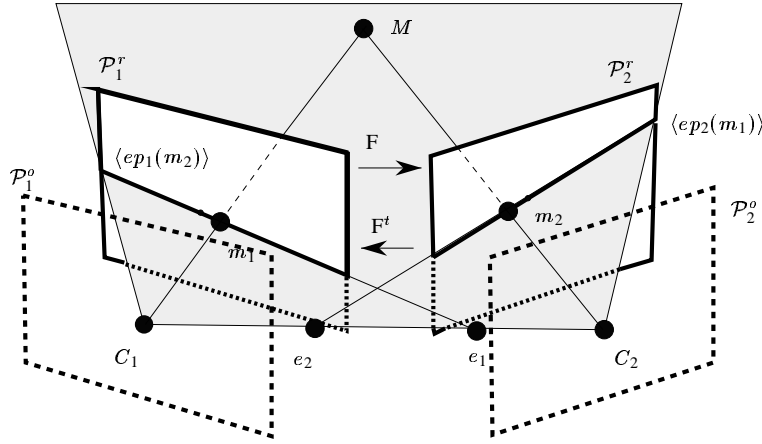


Figure 1: The epipolar geometry: notations.

3 Image properties of planes

Let us consider the projections m_1^k, m_2^k on the two cameras of points M^k that lie on a plane Π in space. It is well known that there exists a homography, i.e. a linear transformation of the projective plane, that maps each point m_1^k onto its correspondent m_2^k . This transformation H can be represented by a 3×3 matrix, that

we denote by \mathbf{H} . The matrix is defined up to a scale factor, so it has $9 - 1 = 8$ degrees of freedom. Thus, a set of four point correspondences in general position (i.e., not three of which are aligned) is sufficient for computing it. Indeed, each point correspondence brings two equations. In other words, two sets of four points that define projective bases of the image planes define only one homographic mapping between the images.

One can show that if the epipolar geometry is known, three point correspondences are sufficient for characterizing \mathbf{H} . A simple reason for that is that the epipoles define a fourth point correspondence [10] which represents the intersection point of the plane Π with the line $\langle C_1, C_2 \rangle$ joining the optical centers (Figure 2). Based on the above result, there exists a unique homography that maps points e_1, m_1^1, m_1^2, m_1^3 onto points e_2, m_2^1, m_2^2, m_2^3 . This is true only when no alignment of three points can be found among the four points in one of the images. In particular, it excludes the case when two of the reference points are aligned with the epipole in one of the images.

We show that the property still holds as long as the three reference points are not aligned. First, we give a geometric proof of this result. Then, we introduce an analytic method that allows computing the homography defined by any three unaligned point correspondences.

3.1 Geometric construction of the homographic mapping

Let us consider four space points $M^1, M^2, M^3, M^4 = M$. If the points are coplanar, the pairwise intersections of the 3D lines $\langle M^a, M^b \rangle$ ($1 \leq a < b \leq 4$) define three other points denoted by N^k , $1 \leq k \leq 3$. The projections of these points in camera i are obtained by intersecting the projected lines $\langle m_i^a, m_i^b \rangle$ pairwise. Since they represent the same point of 3D-space, n_1^k and n_2^k must satisfy the epipolar constraint (Figure 3). Mohr and Faugeras have proved [6] that this last property is necessary and sufficient for the points M^1, M^2, M^3, M to be coplanar.

Let us consider three point correspondences (m_1^k, m_2^k) ($1 \leq k \leq 3$). We show that if the triplets of points in both images are not aligned, then we can compute the image of any point m_1 by the homography of the plane Π passing through M^1, M^2, M^3 . For this we use the coplanarity criterion presented above.

First, we construct the point $n_1^1 = \langle m_1^2, m_1^3 \rangle \cap \langle m_1^1, m_1 \rangle$ (Figure 3), represented by the three-vector $(\mathbf{m}_1^2 \times \mathbf{m}_1^3) \times (\mathbf{m}_1^1 \times \mathbf{m}_1)$. This point is the projection in image 1 of a point N^1 of the plane Π which is aligned with M^2, M^3 . Therefore, the projection

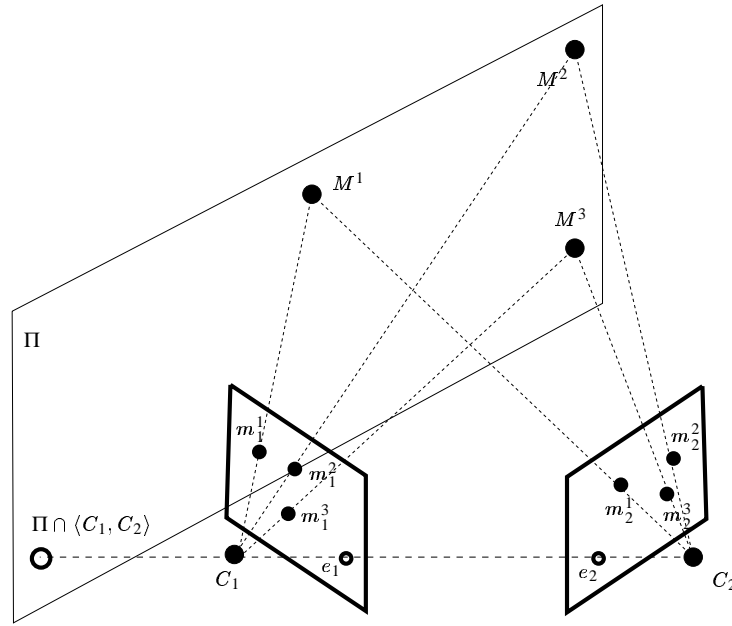


Figure 2: The epipoles represent a fourth point lying on the plane.

of N^1 in the second image, n_2^1 , lies on the line $\langle m_2^2, m_2^3 \rangle$. Since it also lies on the epipolar line of n_1^1 , it is equal to $\langle m_2^2, m_2^3 \rangle \cap \langle ep_2(n_1^1) \rangle$. Since it also lies on the epipolar line of n_1^1 , it is equal to

$$\mathbf{H}\mathbf{m}_1 = \mathbf{F}\mathbf{m}_1 \times (\mathbf{m}_2^1 \times \overbrace{((\mathbf{m}_2^2 \times \mathbf{m}_2^3) \times \mathbf{F}((\mathbf{m}_1^2 \times \mathbf{m}_1^3) \times (\mathbf{m}_1^1 \times \mathbf{m}_1)))}^{\mathbf{n}_2})$$

This can be rewritten as

$$\mathbf{H}\mathbf{m}_1 = \mathbf{F}\mathbf{m}_1 \times \mathbf{G}\mathbf{m}_1 \quad (2)$$

with¹

$$\mathbf{G} = [\mathbf{m}_2^1]_{\times} [\mathbf{m}_2^2 \times \mathbf{m}_2^3]_{\times} \mathbf{F} [\mathbf{m}_1^2 \times \mathbf{m}_1^3]_{\times} [\mathbf{m}_1^1]_{\times}$$

Three analogous expressions can be found, depending on the pair of lines that is chosen to construct n_1 . As long as no three of the four points m_1^1, m_1^2, m_1^3, m_1 lie on the same epipolar line, one can easily find a permutation of the three reference point correspondences with which the method described above can be applied to compute m_2 .

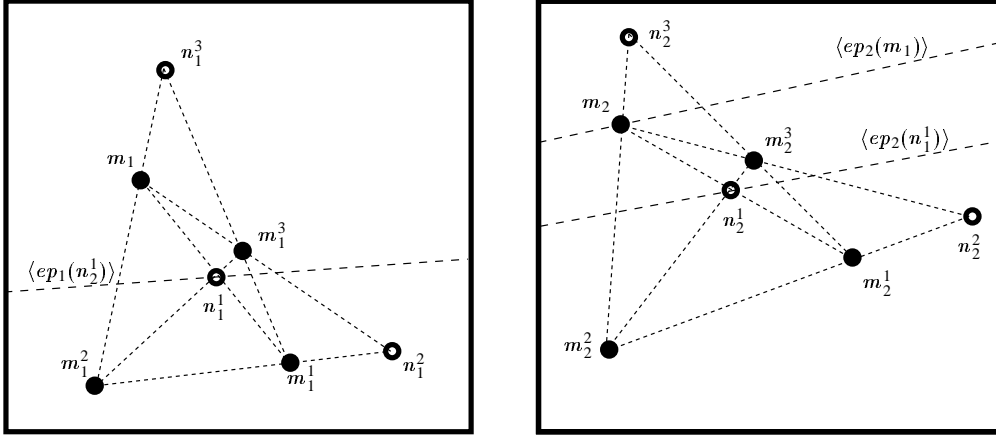
If three of the four points lie on the same epipolar line, since by hypothesis the three reference points m_1^1, m_1^2, m_1^3 are not aligned, we can suppose without loss of generality that the three aligned points are m_1, m_1^1, m_1^2 . In the first image, we choose a point m_1^4 which does not lie on the epipolar line $\langle m_1^1, m_1^2 \rangle$. Using the above geometric method, we determine its image by the homography, m_2^4 . We finally use the point correspondences $(m_1^1, m_2^1), (m_1^2, m_2^2), (m_1^4, m_2^4)$ for computing the image of m_1 by the homography.

This completes the proof that three non-aligned point correspondences define a unique homography between the images. However, though it allows computing the image of any point m_1 by the homography, the method does not yield the homography matrix, and is not very convenient from a practical standpoint. So, in the next section we give an analytic method for computing the homography matrix.

3.2 An analytic approach for computing H

Let us first take into account the fact that the fundamental matrix \mathbf{F} is known. From (1) it follows that for each point m_1 , the corresponding point $\mathbf{H}\mathbf{m}_1$ should lie on the

¹ $[\mathbf{x}]_{\times}$ designs the linear operator defined by: $\forall \mathbf{y}, [\mathbf{x}]_{\times} \mathbf{y} = \mathbf{x} \times \mathbf{y}$

Figure 3: Coplanarity test and construction of $\mathbf{H}(m)$.

epipolar line $\langle ep_2(m_1) \rangle$. Thus,

$$\forall \mathbf{m}_1, \quad \mathbf{m}_1^t \mathbf{F}^t \mathbf{H} \mathbf{m}_1 = 0$$

A necessary and sufficient condition, as remarked in [2], is that

$$\mathbf{F}^t \mathbf{H} + \mathbf{H}^t \mathbf{F} = 0 \quad (3)$$

This equation yields 6 homogeneous equations (the left side is trivially a symmetric matrix). To compute \mathbf{H} we need three other equations.

Let us consider two points m_1, m_2 in homographic correspondence:

$$\mathbf{H} \mathbf{m}_1 \times \mathbf{m}_2 = 0 \quad (4)$$

We show that when (1) and (3) are verified, then (4) is equivalent to the following scalar equation:

$$[\mathbf{m}_2, \mathbf{F} \mathbf{m}_1, \mathbf{H} \mathbf{m}_1] = 0 \quad (5)$$

The implication (4) \Rightarrow (5) is trivial. Reciprocally, if (5) is true, then one can check in a straightforward manner that $\mathbf{H} \mathbf{m}_1 \times \mathbf{m}_2$ is orthogonal to the three vectors $(\mathbf{m}_2, \mathbf{F} \mathbf{m}_1, \mathbf{m}_2 \times \mathbf{F} \mathbf{m}_1)$, which constitute an orthogonal basis of \mathbf{R}^3 . So it is equal to zero.

As a result, each point correspondence holds for exactly one scalar equation on the coefficients of \mathbf{H} . Thus, in the general case, 3 pairs of points are sufficient for computing the homography matrix.

It is important to notice that the system of 9 homogeneous equations (6 for the fundamental matrix, 3 from the point correspondences) represents all the geometric constraints imposed by the fundamental matrix and by the three point correspondences. In the previous paragraph, we have proved that as soon as the points are not aligned, these constraints determine a unique homography. This means that the kernel of the homogeneous system has dimension one.

We end up with a method for computing the homography matrix from $N \geq 3$ point correspondences on the same plane Π , by simply solving with a standard linear least-square technique the following system of $N + 6$ homogeneous equations:

$$\begin{cases} \mathbf{F}^t \mathbf{H} + \mathbf{H}^t \mathbf{F} & = \mathbf{0} \\ [\mathbf{m}_2^1, \mathbf{H} \mathbf{m}_1^1, \mathbf{F} \mathbf{m}_1^1] & = 0 \\ & \vdots \\ [\mathbf{m}_2^N, \mathbf{H} \mathbf{m}_1^N, \mathbf{F} \mathbf{m}_1^N] & = 0 \end{cases}$$

4 Recovering geometric properties of the 3D scene

Based on the homographic correspondences induced by planes in space, we describe three geometric properties that are characteristic of 3D scenes and can be computed from point correspondences under the weak calibration assumption.

4.1 Intersection of lines and planes in space

Let us now consider a line (D) in space, detected by two point correspondences $(m_1, m_2), (m'_1, m'_2)$ which represent two 3D points M, M' that do not both lie on the plane Π . We would like to compute the projections n_1, n_2 of the intersection point N of the line and the plane.

We use a simple geometric argument in order to compute n_2 : On the one hand, N lies on the line $\langle M, M' \rangle$, so n_2 lies on the image line $\langle m_2, m'_2 \rangle$. On the other hand, N lies on the line $\langle D \rangle$ on the plane Π , whose projection in image 1 is $\langle m_1, m'_1 \rangle$ (Figure 4). Since they lie on the plane Π , the points of $\langle D \rangle$ that project in image 1

onto m_1, m'_1 will project in image 2 onto $\mathbf{H}(m_1), \mathbf{H}(m'_1)$. So, the projection of $\langle D \rangle$ in the right image is represented by the vector $\mathbf{H}m_1 \times \mathbf{H}m'_1$. As a consequence, n_2 belongs to the intersection of the two lines $\langle m_2, m'_2 \rangle, \langle \mathbf{H}(m_1), \mathbf{H}(m'_1) \rangle$, and

$$\mathbf{n}_2 = (\mathbf{m}_2 \times \mathbf{m}'_2) \times (\mathbf{H}\mathbf{m}_1 \times \mathbf{H}\mathbf{m}'_1) \quad (6)$$

To compute \mathbf{n}_1 , we can either apply the same construction in image 1, or obtain it directly from $\mathbf{n}_1 = \mathbf{H}^{-1}\mathbf{n}_2$.

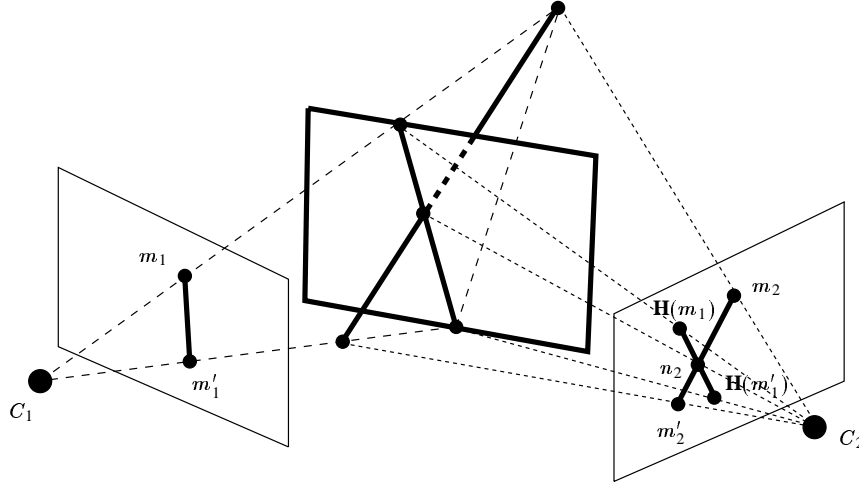


Figure 4: Computation of the projection in the second image of the intersection between a line and a plane (see text).

Remark: \mathbf{H} is compatible with the epipolar geometry (i.e. it satisfies (3)), so points m_2 and $\mathbf{H}(m_1)$ (resp. m'_2 and $\mathbf{H}(m'_1)$) lie on a same epipolar line. Since M or M' does not belong to the plane Π , we have $m_2 \neq \mathbf{H}(m_1)$ or $m'_2 \neq \mathbf{H}(m'_1)$. A direct consequence is that lines $\langle m_2, m'_2 \rangle, \langle \mathbf{H}(m_1), \mathbf{H}(m'_1) \rangle$ are equal if and only if $\langle m_2, m'_2 \rangle$ (or $\langle m_1, m'_1 \rangle$) is an epipolar line. In this case, we can still compute the projections of N , but the construction is slightly more complicated. We choose a point p_1 not on $\langle m_1, m'_1 \rangle$, and a point p_2 on its epipolar line in image 2, distinct from $\mathbf{H}(p_1)$. Points p_1, p_2 represent a space point P that lies neither on the epipolar plane

defined by m_1, m'_1 , nor on the plane Π . Using the method describes above we can compute the projections of $P_\Pi = \langle P, M \rangle \cap \Pi$ and $P'_\Pi = \langle P, M' \rangle \cap \Pi$. By construction, the two space lines $\langle M, M' \rangle$ are coplanar, distinct, and their intersection point lies on Π (Figure 5), so it is equal to N . As a consequence, by pairwise intersecting the corresponding image lines we obtain the projections of the point N .

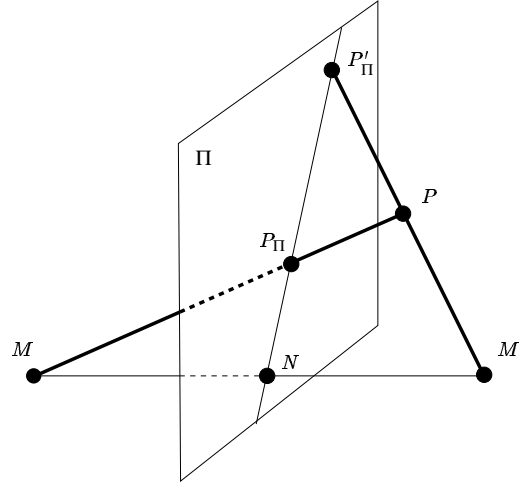


Figure 5: Computation of the intersection between a line and a plane in the degenerated case (see text).

4.2 Visibility with respect to a camera

We call *visibility* the constraint that all 3D points imaged by a physical camera must lie in the half-space bounded by its optical plane. Points that lie “behind” the optical plane are called *invisible*.

Since \mathcal{P}^o is the locus of three-dimensional points that project in the image onto the line at infinity (i.e. the set of all points at infinity), the image of an object that is completely visible or completely invisible contains no point at infinity (Figure 6).

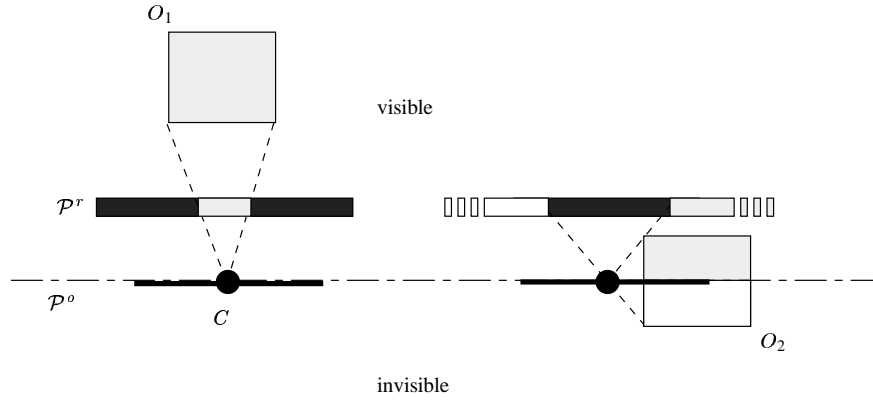


Figure 6: Contrary to O_1 , object O_2 is partially visible and partially invisible. Its image contains at least a point at infinity.

4.3 Position with respect to a plane

Under the same hypotheses as in paragraph 4.1, we now assume that points M and M' are visible for both cameras. Then the line segment $[M, M']$ intersects none of the optical planes $\mathcal{P}_1^o, \mathcal{P}_2^o$, and its projection on retina i contains no point at infinity: It is the euclidean line segment $[m_i, m'_i]$. As a consequence,

$$N \in [M, M'] \Leftrightarrow n_1 \in [m_1, m'_1] \Leftrightarrow n_2 \in [m_2, m'_2]$$

We have shown above how to compute the position of n_i on line $\langle m_i, m'_i \rangle$. Thus, we can test whether the 3D points M, M' lie in the same half-space limited by Π , by checking *directly in the image* whether n_1 lies between m_1 and m'_1 .

Remark: This property relies on the fact that segment $[M, M']$ does not intersect the optical planes $\mathcal{P}_1^o, \mathcal{P}_2^o$. Consequently, from a purely geometric standpoint, it remains valid if M and M' are invisible for both cameras.

4.4 Convex hull of a three-dimensional object

Let us now show how the previous properties can be used to determine the convex hull of a set of *visible* points, given their images. We consider a set of space points

$\mathcal{E} = \{M^i, 1 \leq i \leq N\}$. A plane containing at least one point of \mathcal{E} is said to be extremal if and only if all the points of \mathcal{E} lie in the same half-space limited by this plane. A subset of an extremal plane is also said to be extremal.

From these definitions, the convex hull of \mathcal{E} (denoted by $\mathcal{C}(\mathcal{E})$) is the union of the extremal polygons whose vertices belong to \mathcal{E} [9]. It is also the union of extremal triangles whose vertices belong to \mathcal{E} .

A “brute-force” algorithm: Using the result of section 4.3, we can easily detect if a triangle defined by its three vertices is extremal. Indeed, we test whether all the other points of \mathcal{E} lie on the same side of the plane of the triangle. Points are compared pairwise, and the complete test is performed in $O(N)$ operations. We end up with a very simple algorithm, for computing the convex hull of \mathcal{E} given the point matches in weakly calibrated images: For each possible point triplet of \mathcal{E} , test if the points define an extremal triangle. The complexity of this “brute force” approach is obviously $O(N^4)$.

The “gift-wrapping” algorithm: Other techniques with better complexity have been designed for computing the convex hull of a set of points in three dimensions. For instance, the “gift-wrapping” technique [9] computes the convex hull in $O(N^2)$ operations. It assumes that all the facets of the convex hull are triangles. The algorithm is two-fold: First, in a bootstrapping stage, one facet of the convex hull is built in linear time. Then, the convex hull is built incrementally, with complexity $O(N^2)$. The second stage can be visualized as wrapping a sheet, facet by facet, around the points of \mathcal{E} . More precisely, given a facet $F = \{M^a, M^b, M^c\}$ of the convex hull, the facet F' adjacent to F and containing edge $[M^a, M^b]$ is the one that maximizes the angle with F around this edge (Figure 7).

We now show how to apply the same technique to the weakly-calibrated case.

Bootstrapping: We first build a facet of the convex hull $\mathcal{C}(\mathcal{E})$, with quadratic complexity:

- In one image, we determine one edge a of the two-dimensional convex hull of the projected points. By definition, the other image points all lie on the same side of a . As a consequence, the three-dimensional points all lie on the same

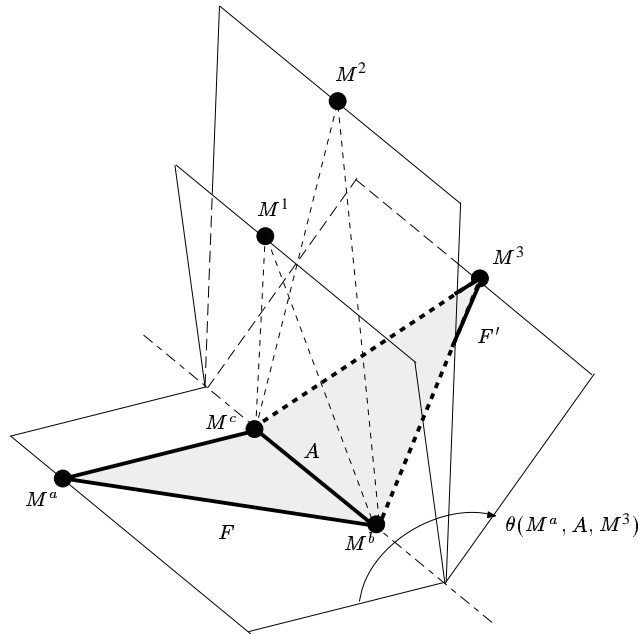


Figure 7: Facet F' is defined by point M^3 , which maximizes $\theta(M^a, A, M)$ for $M \in \mathcal{E} \setminus \{M^a, M^b, M^c\} = \{M^1, M^2, M^3\}$.

side of the plane passing through the corresponding three-dimensional edge A and the optical center of the camera: A is extremal, so it belongs to $\mathcal{C}(\mathcal{E})$.

- For each point M of $\mathcal{E} \setminus \{A\}$, test whether the plane defined by A and M is extremal. We apply the “brute-force” technique to perform this test.

As a result, we have built one facet of the convex hull with complexity $O(N^2)$.

Gift-wrapping: Let us consider a facet $\{M^a, M^b, M^c\}$ that belongs to the convex hull, and denote by A the edge $[M^b, M^c]$. For a given point M^i , we denote by $\theta(M^a, A, M^i)$ the angle between the facets $\{M^b, M^c, M^i\}$ and $\{M^a, M^b, M^c\}$. The basic stage of the gift-wrapping process consists of finding the point $M^i \in \mathcal{E} \setminus \{M^a, M^b, M^c\}$ that maximizes $\theta(M^a, A, M^i)$ (Figure 7).

Under weak calibration, we do not have a Euclidean description of the set of points. Therefore, we cannot compute the angles $\theta(M^a, A, M^i)$. However, given two points $(M^1, M^2) \in \mathcal{E} \setminus \{M^a, M^b, M^c\}$, it is possible to compare the angles $\theta(M^a, A, M^1)$ and $\theta(M^a, A, M^2)$ without computing them explicitly. Indeed, $\theta(M^a, A, M^1)$ is smaller than $\theta(M^a, A, M^2)$ if and only if M^a and M^2 lie on opposite sides of the plane defined by A and M^1 (Figure 8). This property can be tested using the result of section 4.3. This completes the description of a version of the gift-wrapping method for computing in quadratic time the convex hull of a set of points in space given by their projections in weakly-calibrated images.

5 Experimental issues

In this section we show some experiments conducted on point correspondences extracted from a pair of real images.

We first computed the fundamental matrix of the pair of images using the algorithm presented in [11]. This algorithm computes point correspondences in the images based on a correlation criterion, and extracts the fundamental matrix of the pair of cameras from these point matches. Please note that the system of cameras is *weakly calibrated*, i.e. the only geometric knowledge we have is the epipolar transformation between the cameras, represented by the fundamental matrix.

Then, we manually selected some sets of point correspondences and computed their convex hulls. The images and the set of point correspondences that are used are

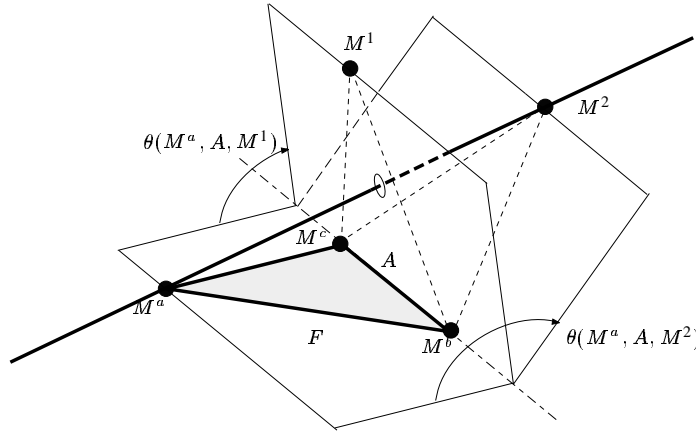


Figure 8: $\theta(M^a, A, M^1)$ is smaller than $\theta(M^a, A, M^2)$ if and only if points M^a and M^2 are on opposite sides of the plane (M^b, M^c, M^1) .

represented in Figure 9. The two cross-fusion stereograms of Figure 10 represent the convex hulls computed for the three corresponding sets of points separately (top row), and the convex hull of the total set of points (bottom row). For clarity, we represent edges that are hidden in the first image with thin lines, and the other edges with thick lines. The images in Figure 10 are slightly different from the original ones: we have rectified them so that the reader can perform cross-fusion more easily. The rectification transformations are determined directly from the fundamental matrix through a technique analogous to the one described in [3], and have also been applied to the points and the convex hulls.

6 Conclusions and perspectives

We have shown that under the “weak calibration” hypothesis in which only the epipolar geometry of a pair of images is known, we can extract from the images very useful information about relative positions of points and planes, without using any three-dimensional description of the objects. Using these properties, we have designed two algorithms for determining the convex hull of a set of points in space,

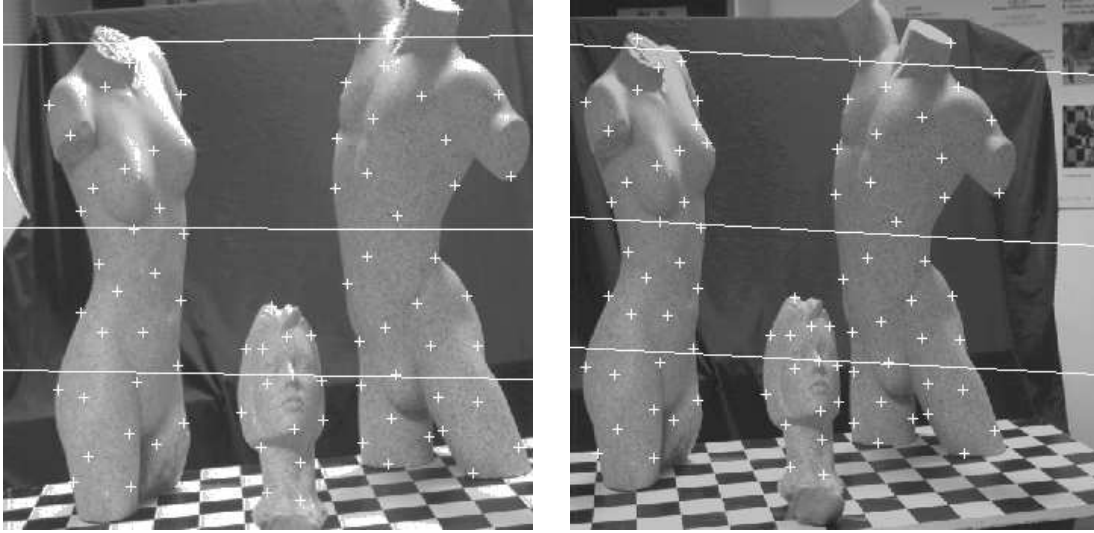


Figure 9: The pair of weakly-calibrated images, point correspondences, and a set of epipolar lines

represented by their projections in two weakly-calibrated images. Results obtained on real data show the efficiency of these techniques.

We believe that it is essential to develop elements of geometric reasoning under the weak calibration assumption. First, it is a step toward robustness. Indeed, when solving a problem by computing things directly “in the images”, we avoid introducing intermediate variables such as three-dimensional Euclidean coordinates. These variables are not necessary for solving the problem, and their computation can introduce some additional noise. Second, from a practical standpoint, we avoid the painful step of camera calibration. A robotic system equipped with a set of cameras can perform weak-calibration on images of its environment, without requiring to use a special calibration object. In this respect, geometric reasoning under weak calibration constitutes a major step toward the development of autonomous robotic systems.